

HOW A PERSONAL DOCUMENT'S INTENDED USE OR PURPOSE AFFECTS ITS CLASSIFICATION IN AN OFFICE

Barbara H. Kwasnik
School of Information Studies
Center for Science and Technology
Syracuse University
Syracuse, N.Y. 13244-4100 USA
bkwasnik@suvvm.bitnet

1. ABSTRACT

This paper reports on one of the findings of a larger case study that attempts to describe how people organize documents in their own offices. In that study, several dimensions along which people make classificatory decisions were identified. Of these, the use to which a document is put emerged as a strong determiner of that document's classification. The method of analysis is reviewed, and examples of different kinds of uses are presented, demonstrating that it is possible to describe a wide variety of specific instances using a closed set of descriptors. The suggestion is made that, in designing systems for organizing materials, it might be advantageous to incorporate information about contextual variables, such as use, since these seem to be particularly important in classification decisions made within personal environments.

2. INTRODUCTION

The work described in this paper is part of a study whose aim is to describe how people, in everyday situations, make classificatory choices (KWASNIK, 1989). The assumption is that these choices are never made in a vacuum, but rather in a context or situation (MISCHLER, 1979). The implicit point of view is that even though objects themselves provide some constraints on classification choices, classification is, overall and above all, person- and situation-centered, and not object-centered.

Thus, the focus of the study is on situations and on methods of eliciting data that will provide a description of how people make classificatory choices within a given situation. It is a case study describing the process by which people organize and classify documents in their own offices as an example of a typical classificatory situation. Put another way, it is a study of "situated meaning," which assumes that people actively create meanings in the context of a given situation (DERVIN, 1983). A goal of the study is to describe those dimensions of a person's situation that are significant with respect to classificatory behavior.

The observations made in such related studies as those of CASE (1986) and MALONE (1983) are extended and amplified by a more detailed analysis of the data and (with respect to the study by Case) by a slightly different method of data collection. The main focus is to describe and analyze situations and contexts, rather than the individuals and the objects themselves.

3. METHODOLOGY

The main concerns addressed by the methodology were: 1) to collect data in as natural a setting as possible so that context could be observed and incorporated into the description; 2) to take into account the difficulties of collecting verbal data on what is essentially a cognitive process (ERICSSON & SIMON, 1976); and 3) to ensure that subjects be allowed to generate their own descriptions, labels, and relationships, rather than respond to descriptions, labels, and relationships decided upon a priori by the researcher (FRAKE, 1969; TYLER, 1969).

4. SCOPE

The purpose of the study as a whole was to build up a description of the objects defined as documents found in a person's office, described in the person's own terms, and the relationships among those documents; the circumstances impinging on classification choices; the important dimensions along which objects are classified; and the rules used in guiding these classification choices.

This paper will report on only one of the important dimensions along which classificatory decisions are made, namely the intended use or purpose of a document. This facet or dimension deals with the uses to which something had been, is being, or will be put.

5. PROCEDURES

Eight university faculty members, men and women from a variety of disciplines and of various academic ranks, were asked to describe their own offices in terms of the organization of what each of them defined as documents. Each subject was also asked to sort a day's mail, simulating as closely as possible the usual way in which this task is done, but "thinking out loud" and giving as much detail as necessary to describe the process to another person. The initial data sets thus included interview protocols (for retrospective data), thinking-out-loud protocols (for task concurrent data), and notes on observations made by the researcher.

Following the data analysis, four of the eight subjects were interviewed once again. During this session, the researcher used the results of data analysis as a guide and tried to sort and classify each subject's mail and several days' worth of accumulated documents following the same rules and criteria that he or she might have used. The subject was asked to comment on the accuracy of the decisions made by the researcher. This session was in part a validation of the analysis, and in part an additional method of data collection.

6. DATA ANALYSIS

6.1 Identifying the Dimension of Use

The protocols from the first two interviews ("grand tour" and mailsort) were transcribed verbatim. These transcripts were first analyzed for arrays of objects, i.e., "documents." A document was identified when it was labelled or named or when it was referred to by a pronoun or by a general noun, such as, for example, "stuff." Documents could be identified by nouns or noun phrases.

Examples of objects/documents in a person's office are:

"a program for computing the location of comets"
 "books on ethics"
 "a note from somebody about lunch"
 "my fall semester schedule of classes"

The labels for these documents were used as a focus or starting point for further analysis. Documents were often defined and described by means of modifying phrases. In talking about a document, or group of documents, the subject often described the circumstances of the classificatory decisions pertaining to it, or the reasons for its placement or grouping. These modifiers and explaining phrases were summarized by brief terms or labels, defined into codes, and then used to describe other instances in which such a description might apply.

For example, the following two instances of classificatory decisions:

[On the top shelf are books that are very seldom used.]
 [Correspondence I must deal with immediately goes into my briefcase.]

can both be described by the same set of codes:

LOCATION
 on the top shelf
 into my briefcase

FORM
 books
 correspondence

TIME
 very seldom
 immediately

USE
 used
 deal with

All of the cases were analyzed in this way, that is, starting with the identification of the document and proceeding to a description of the dimensions along which classificatory decisions pertaining to each document were made. This analysis yielded an inventory of documents, an inventory of dimensions or criteria that people mentioned as part of the classificatory process, as well as an indication of how frequently each dimension was invoked with respect to the classification of documents.

6.2 Classification of Dimensions

Once the entire corpus had been coded, it was possible to merge and rearrange the categories so that extremely fine levels of distinction that accounted for a very small proportion of the data were collapsed into more inclusive categories. This yielded 35 categories, which could then be arranged into seven even broader groups.

The seven descriptive coding groups that represent dimensions used in making classificatory decisions are: Situation Attributes, such as source, use, circumstance, and access; Document Attributes, such as author, topic, and form; Disposition, such as discard, keep, postpone; Order/Scheme, such as group, separate, and arrange; Time, such as continuation, duration, and currency; Value, such as importance, interest, and confidentiality; and Cognitive State, such as "don't know," and "want to remember." Individual classification decisions could be multiply coded.

6.3 Identifying the Relative Importance and Citation Order of the Dimensions

Once the data had been coded it was possible to compute the frequency with which dimensions were invoked in making classificatory decisions, but frequency alone does not indicate the relative importance of the dimension to the decision, nor does it indicate the order in which it should be invoked. For example, in:

"These are things I use for teaching"

the dimension that defines the category is the use of the documents (regardless of form, topic, physical attributes, and so on). On the other hand, in:

"These are books I use for teaching"

both form (book) and use (teaching) define the category together. Finally, in:

"These are current materials — a book to review, some recent correspondence, and an agenda for tomorrow's meeting"

the defining dimension is time (currency), while form, use, and purpose are mentioned but do not take precedence over time.

Thus, we see that dimensions along which classificatory decisions are made can be defining dimensions, can share the function of definition with another dimension or other dimensions, or can be identified but not be defining.

Furthermore, it is important to describe the order in which dimensions are invoked. For example, if exams in a particular office are filed in two places: in one place if the exams are for courses taught previously, and in another if they are for courses taught currently, then in order to know where a document labelled "exam" is to be put, it is first necessary to identify its form and its purpose, and then to identify the course to which it belongs, and then to determine whether the course is currently taught or not, and then to locate it in the proper folder in the proper drawer. Knowing the sequence in which dimensions are invoked is crucial to an accurate description of the decision.

Towards this end, each identified classificatory decision in all the cases was further analyzed to discover and describe not only the frequency of classificatory dimensions, but also the order in which they are invoked and the degree to which they are defining dimensions. Each classificatory decision was rephrased as a rule. The following example is typical:

IF:	file folders	FORM
	for courses	TOPIC
	taught	USE
	last year	TIME
AND		
IF:	not used	USE
	on daily basis	TIME
THEN:		
	place	LOCATE
	on desk	CONTAINER
	in a pile	ORDER

While much of the information about the classificatory decision remains implicit, the analysis does give a rough picture of the relative importance of the dimensions.

7.RESULTS

7.1 Frequency of Use as a Dimension

Situation Attributes accounted for approximately 35 percent of all the descriptive dimensions of classificatory decisions and were the most frequently cited overall. Of the Situation Attributes, use was the most frequently cited, and second only to form overall. Use was not only frequently cited but was also important in defining categories, that is, in those situations where a category could be defined by several dimensions, use was either the defining or codefining variable.

A classificatory decision was labelled as USE when the use or purpose to which a document had been, was, or would be put was a criterion of its categorization. As with all the dimensions, but with this one in particular, it was often difficult to disentangle a document's use or purpose from other dimensions, such as topic, circumstance, and form. For example, a lecture or a project has the combined, if implicit, dimensions of FORM, TOPIC, and USE. That is, a lecture has a specific format if it is written out; it is about something; and presumably it is used in the process of

teaching. Similarly, if a person says, "These are all my things from the tenure process," we can infer that the documents were both about and for use during the tenure process. In such cases the text was multiply coded. In general, words such as "project," "process," "research," and so on, were taken to specify use even if it was not explicitly so stated.

7.2 Kinds of Uses

Subjects identified many instances and kinds of use that impinge on the classification of documents. Fig. 1 lists some of the uses identified by the participants. They are listed in no particular order and are meant to suggest the variety of activities that come under this coding category.

Fig.1. SOME USES THAT AFFECT THE CLASSIFICATION OF DOCUMENTS

work on a report	keep up with something
send a letter	calculate an inclination
do graphics routines	reproduce a print
fill out a form	study a subject
apply for a grant	list output to the screen
attend to something	adopt a book for courses
publish a book	take minutes at a meeting

Most of these specific uses and purposes can be roughly summarized by the broad, not mutually exclusive categories of document uses outlined in Fig. 2.

Fig.2. BROAD CATEGORIES OF DOCUMENT USE AFFECTING CLASSIFICATORY DECISIONS

<p>TWO USES TOGETHER</p> <p>whenever I was working on one kind of insurance I was almost always working on more than one kind of insurance</p> <p>USE MODIFIED BY TOPIC</p> <p>everything that has to do with fundraising from the private sector</p> <p>USE UNDER CERTAIN CIRCUMSTANCES</p> <p>letters that I write to the newspaper when I'm particularly pissed off</p> <p>USE IN A CERTAIN PLACE</p> <p>stuff I read at home</p> <p>USE DEFINED BY TIME</p> <p>PRESENT</p> <p>the project I'm involved in</p> <p>PAST</p> <p>a recent paper I've just finished</p> <p>FUTURE</p> <p>that file will eventually get modified and expanded</p> <p>OCCASIONAL</p> <p>a directory I sometimes need to refer to</p>
--

IMMEDIATE

a whole bunch of stuff that I have to respond to right away

HABITUAL

usually I sort my mail right down in the mail room

USE THAT ACCOMPLISHES SOME TASK

a program to calculate an inclination

AN ENABLING USE

that's just my test directory so I can do graphics routines

A REQUIREMENT OR REQUEST

a student may want me to write a letter of recommendation

NO USE

I never use this anymore

USE IS IMPLIED

all the preparation for accreditation
an active file

It was, therefore, possible to describe a wide variety of specific situations by a closed group of descriptors or codes.

7.3 Use as a Criterial Dimension

Although people identify many dimensions or aspects of objects, not all have the same weight in making classificatory decisions. Use has emerged as an important or criterial dimension. In other words, the use to which something is put is often the basic level at which it will be classified. For instance, although a person may distinguish between books on various topics, books acquired at different times, and books of different formats and sizes, all of them may be physically as well as cognitively grouped as "books used in teaching Anthro 101."

The dimension of use or purpose for classification of documents has an especially strong correlation with other dimensions, particularly those suggested by MALONE (1983). In this study these dimensions were called "access," "time," and "remembering." For example:

This is a three-year budget for a proposal that I'm submitting that I haven't gotten around to typing yet, but I don't want to file it 'cause I don't want to forget about it, so it sits here.

A document's intended use or purpose is often the first classificatory rule invoked. Documents may be further divided and organized, but the first cut is frequently determined by use or purpose, or use in combination with another dimension.

8. DISCUSSION AND CONCLUSIONS

In their work on classification structures, ROSCH (1975) and others offer the concept of the cognitive reference point. These reference points are cues that act as anchoring points in making classificatory decisions. The research indicates that we choose

categories, at least for concrete objects, that have the most usefulness for the least cognitive effort. While their research refers only to concrete objects such as chairs, birds, cups, and so on, we might consider whether these concepts might not be extended to non-concrete aspects of objects as well, aspects such as use, purpose, urgency, and importance. In other words, to what extent do dimensions of classification such as purpose, use, time, form, topic, and so forth, act as the cognitive reference point for a category?

The data from this study suggest that the dimension of use has great power in determining the classification of a document in the everyday situation of a personal office. The implications for system design are obvious. In addition to modelling the objects that constitute a system, perhaps it would be advantageous to explore ways of modelling typical contexts, circumstances, uses and purposes as well — criteria that seem to be particularly important in classification decisions made within personal information environments.

9. REFERENCES

- CASE, D.O. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science* 11 (1986).
- DERVIN, B. Information as a user construct: the relevance of perceived information needs to synthesis and interpretation. In: Ward, S.A. & Reed, L.J. (Eds.). *Knowledge Structure and Use: Perspectives on Synthesis and Interpretation*. Philadelphia, Pa.: Temple University Press, 1983: 153-83.
- ERICSSON, K.A. & SIMON, H.A. Verbal reports as data. *Psychological Review* 87(3) (May, 1980):215-251.
- FRAKE, C.O. The ethnographic study of cognitive systems. In: Tyler, S. A. (Ed.). *Cognitive Anthropology*. New York, N.Y.: Holt, Rinehart & Winston, 1969.
- KWASNIK, B.H. The influence of context on classificatory behavior. Ph.D. Dissertation, Rutgers University, 1989.
- MALONE, T.W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems* 1(1) (Jan., 1983):99-112.
- MISCHLER, E.G. Meaning in context: is there any other kind? *Harvard Educational Review* 49(1) (Feb., 1979):1-19.
- ROSCH, E. Cognitive reference points. *Cognitive Psychology* 7(4) (1975):532-47.
- SPRADLEY, J.P. *The Ethnographic Interview*. New York, N.Y.: Holt, Rinehart & Winston, 1979.
- TYLER, S.A. (ED.) Introduction. In: *Cognitive Anthropology*. New York, N.Y.: Holt, Rinehart & Winston, 1969.